Project report

Authors: *Tran Huong Giang Pham*

Master of Computer Science.

Emails: t.pham1@studenti.unipi.it

1 Time series and time series forecasting

Time series is a series of data points indexed in time order. These data points are used to build a model to predict future values. This process is called time series forecasting. The data points (or observations) are often assumed to be available at a discrete and equispaced interval of time. Time series has some components:

1. Level: the average value in the time series.
2. Trend: the increasing or decreasing value of the time series.
3. Seasonality: The repeating short term cycle in the time series
4. Noise: Non-expected variability in the observation that cannot explained by the model

There are several methods for time series forecasting and each method is suitable for a time series with different components.

1. Autoregression $AR(p)$ is a model with order p that is defined: $X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \epsilon_t$.
   Autoregression is suitable for non-seasonal and non-trendy data.
2. Moving Average $MV(q)$: is q^{th} moving average model that is defined as:
   $x_t = \mu + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + ... + \theta_q w_{t-q}$
   with w_t is identical, independent and normal distributed variable with the mean 0 and the same variance for every t which is σ_w^2 that represent the white noise errors, μ is the mean of the series and θ_i is the parameter. This model is suitable for time series without trend and seasonality.
3. Autoregressive integrated moving average $ARIMA(p, q, d)$ is a model combine from the autoregressive, integrated and moving average parts. When combine the autoregressive and moving average part we have the model:
   $(1 - \sum_{i=1}^{p\prime} \alpha_i L^i) X_t = (1 + \sum_{i=1}^{q} \theta_i L^i) \epsilon_t$

with L is the lag operator, α and θ are the parameters of autoregressive and moving average model.
Now if the polynomial $(1 - \sum_{i=1}^{p\prime} \alpha_i L^i)$ has a unit root that is a factor $(1 - L)$ of multiplicity d then the above model can be rewrite:
$(1 - \sum_{i=1}^{p} \varphi_i L^i)(1 - L)^d X_t = \delta + (1 + \sum_{i=1}^{q} \theta_i L^i)\epsilon_t$ with $p = p\prime - d$.
This is the $ARIMA(p, d, q)$ model.

4. Seasonal Autoregressive Integrated Moving Average $SARIMA(p, d, q)(P, D, Q)_m$ model is the $ARIMA(p, d, q)$ which is added with a seasonal part $(P, D, Q)_m$ in which P, D, Q have the same meaning as p, d, q in the original model but with the Lag operator of the seasonal period and m is the number of season per year. For example the model $SARIMA(1, 1, 1)(1, 1, 1)_4$ is define as:
$(1 - \varphi_1 L)(1 - \phi_1 L^4)(1 - L)(1 - L^4)X_t = (1 + \theta_1 L)(1 + \Theta_1 L^4)\epsilon_t$.

To experiment the time series forecasting methods, I simulate some ODEs to get the data then split the obtained data into 2 parts: one for building the forecasting model and the other one is to evaluate the model.

2 Experiments

2.1 Lotka-Volterra predator/prey model

Firstly, I tried with the Lotka-Volterra predator/prey model by solver ode23 from Matlab with the initial state is $[20, 20]$,$\alpha = 0.01$, $\beta = 0.02$ and with 100 time steps. The simulation result is shown in Fig. 1.
Clearly, this data has seasonality, so I decide to use methods that can deal with time series with seasonality component. Thus, the method I decided to try is Seasonal autoregressive integrated moving average (SARIMA). Firstly, I used data from 30 first steps to build the forecasting model to predict the next 70 time steps but one by one, after getting a predict result, I add to the series with the new observation and then fit the model again. The predicted the observed value of prey and predator are shown in Fig 2. Then I try with longer prediction.
I use the first 50 time steps as the training data and predict for a long term data of 50 time steps. Fig 3 show the predict result for prey and predator.

2.2 Adding noise to Lotka-Volterra model

I try to add noise from normal distribution with zero mean and standard derivation is 5, the data is show in Fig 4. Then I apply one by one time step predict and long term predict. The one by one step predict result for prey and predator is shown in Fig 5 and the long term predict result is shown in Fig 6.

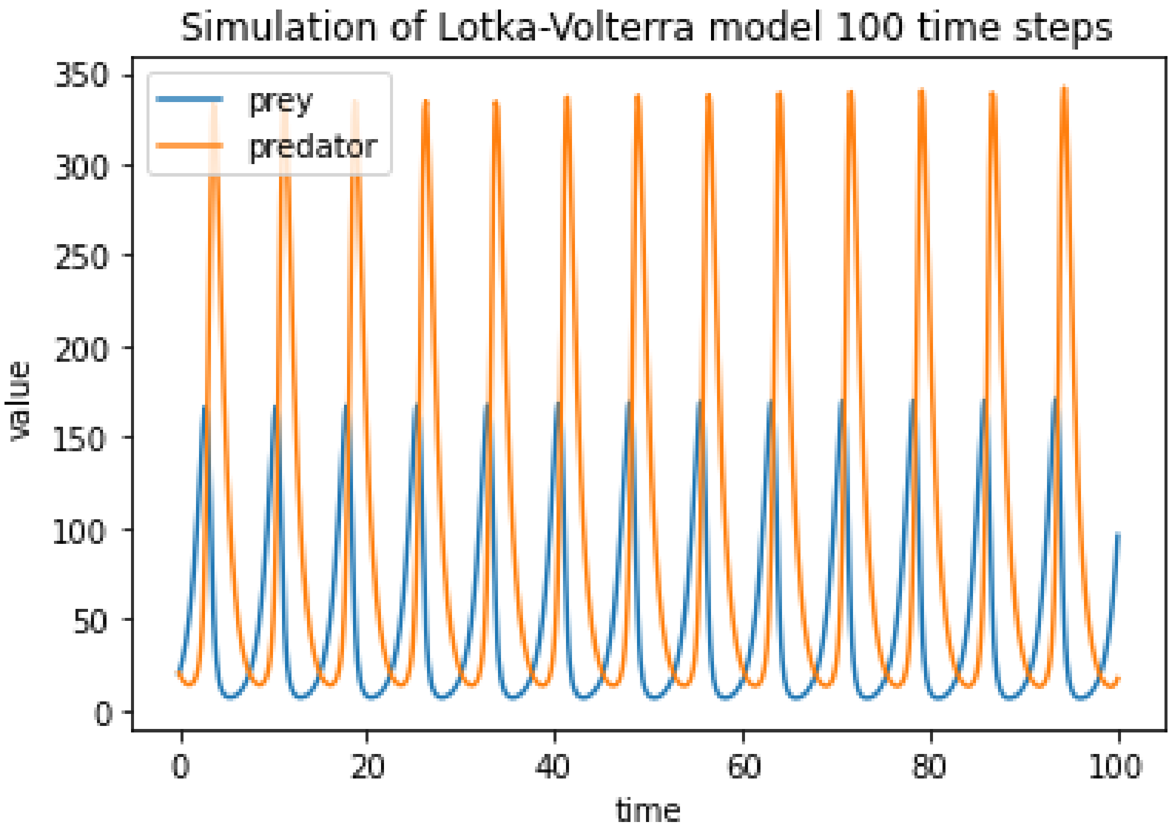


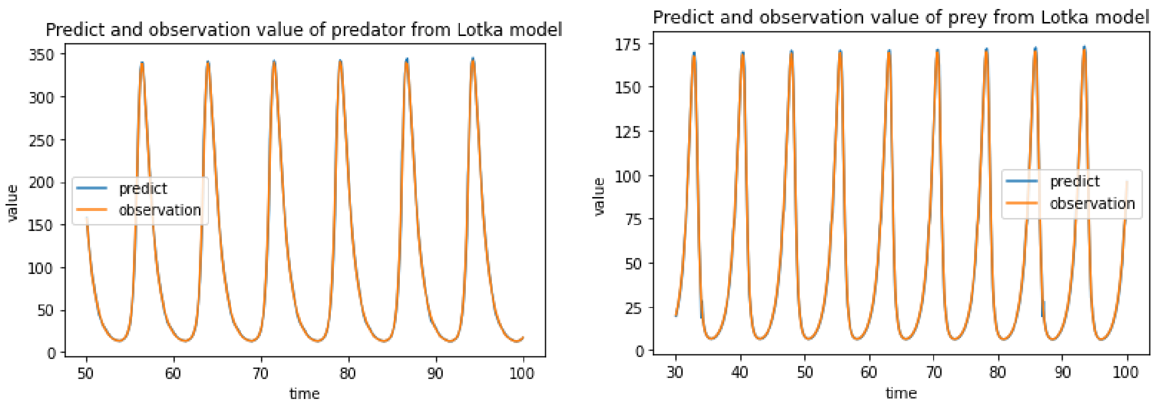
Figure 1: Simulation of Lotka/Volterra model with 100 time steps

2.3 SIR model

Next model I try is SIR model with 100 time steps and $S_0 = 0.99, I_0 = 0.01, R_0 = 0, \beta = 6, \gamma = 2, \mu = 2$. The simulation is show in Fig 7. Then I fit the ARIMA model with 50 first time steps and then predicts for the next 50 times steps. The predict results of S, I, R are shown in Fig 8. Then I try with few initial time steps. I fit the ARIMA model with only 10 time steps and predict for the next 90 time steps. The result is show in Fig 8.

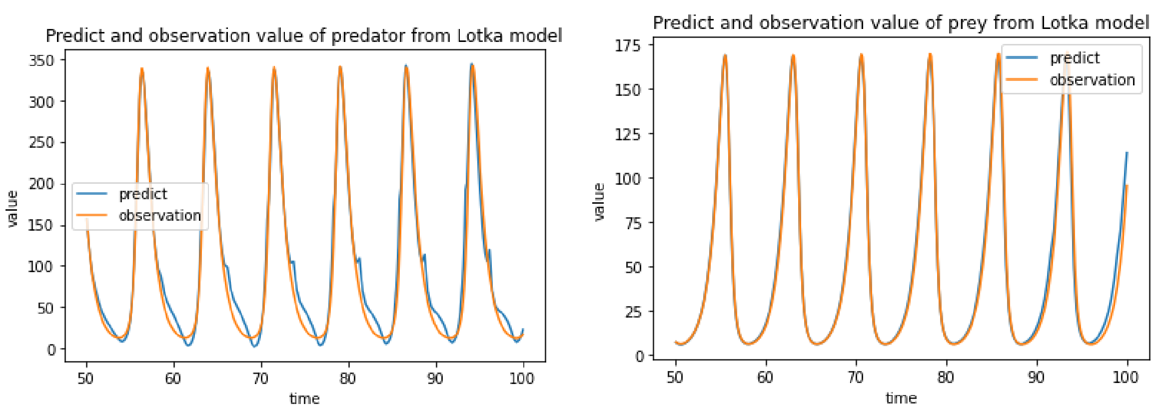
2.4 Logistic equation

The final model I try is the logistic equation with $N_0 = 10$, $r = 4$ and $K = 50$. The simulation result is shown in Fig 10. I use the first 50 data points to fit the model and then predict the next 50 data points. The predict result is shown in Fig 11



(a) (b)

Figure 2: Predict result one by one for (a) predator (b) prey.



(a) (b)

Figure 3: Predict result of 50 time steps for (a) predator (b) prey.

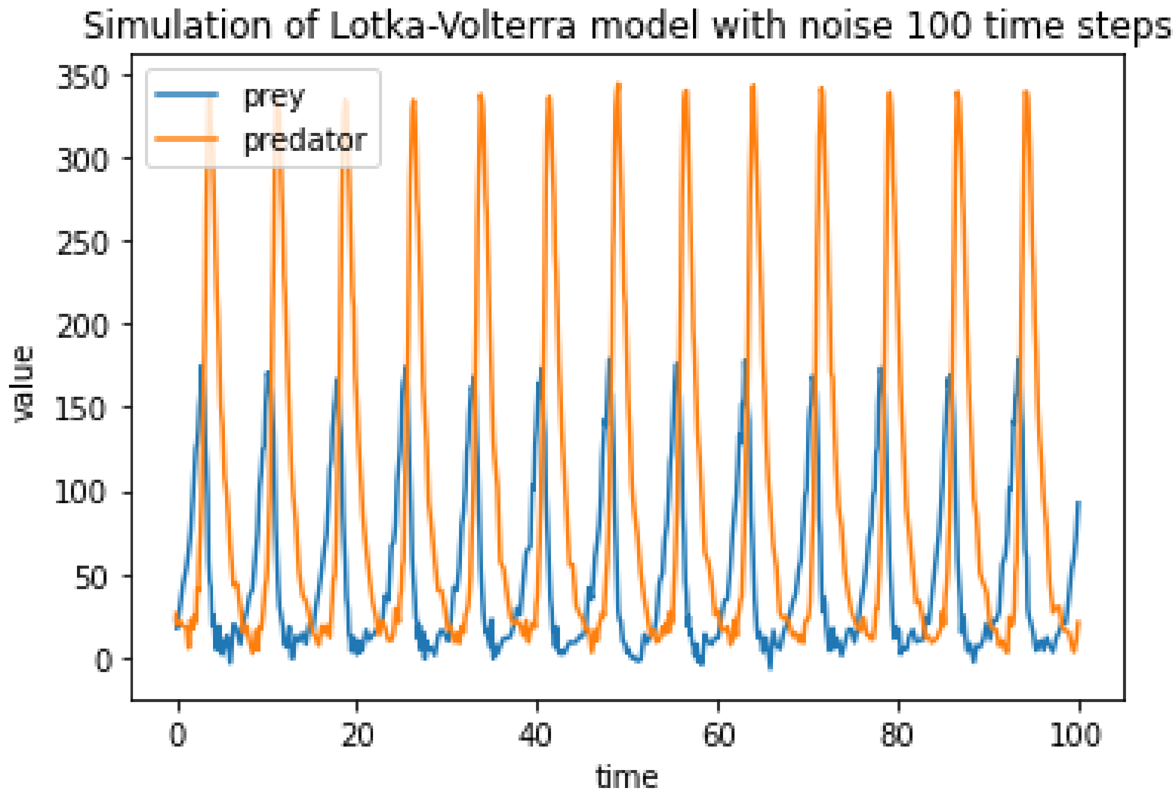
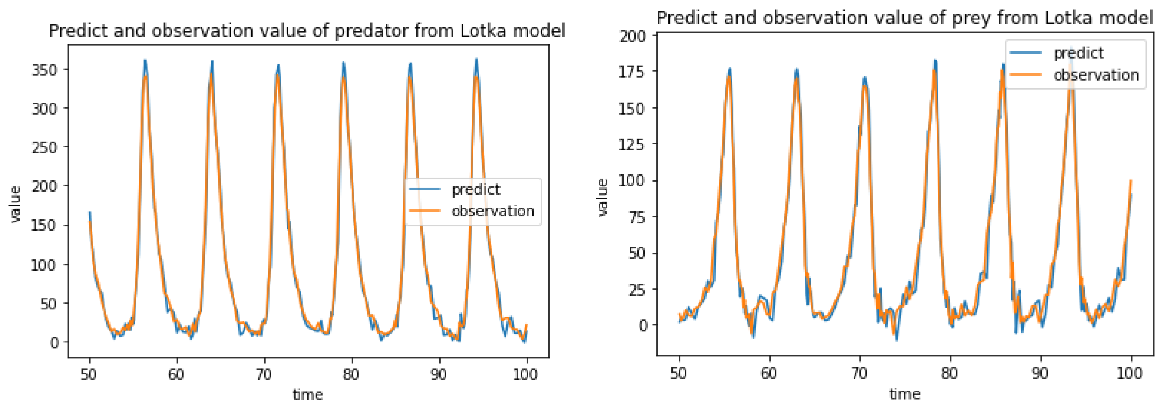


Figure 4: Simulation of noisy Lotka/Volterra model with 100 time steps

3 Conclusion

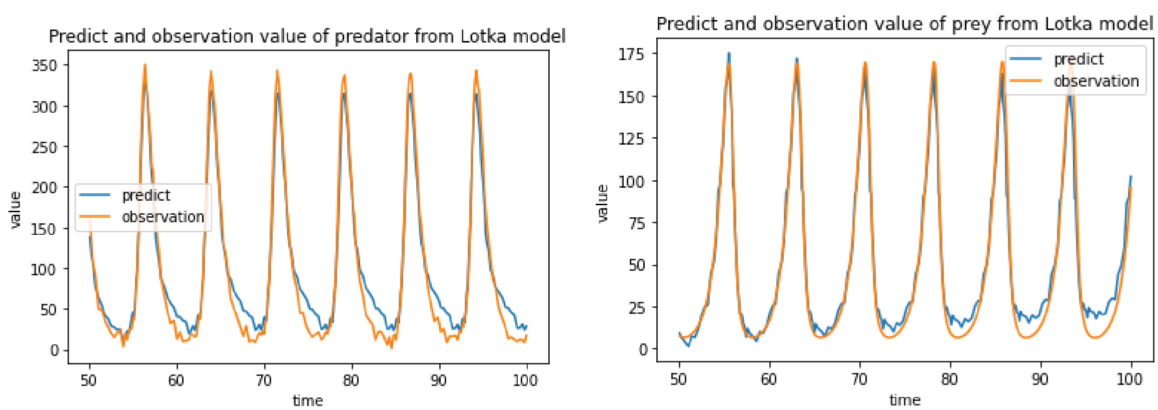
The experiment above leads me to some conclusion.

1. Some time series forecasting methods can be use to predict the future data of a complex model.
2. However, the result gets worse for the further future.
3. To get an decent predict result, at least an enough number of data need to be used to fit the model
4. For the data that fluctuates strongly without any pattern over a short period of time, the result is not good.



(a) (b)

Figure 5: Predict result one by one for (a) predator (b) prey of noisy Lotka-Volterra model.



(a) (b)

Figure 6: Predict result of 50 time steps for (a) predator (b) prey of noisy Lotka-Volterra model.

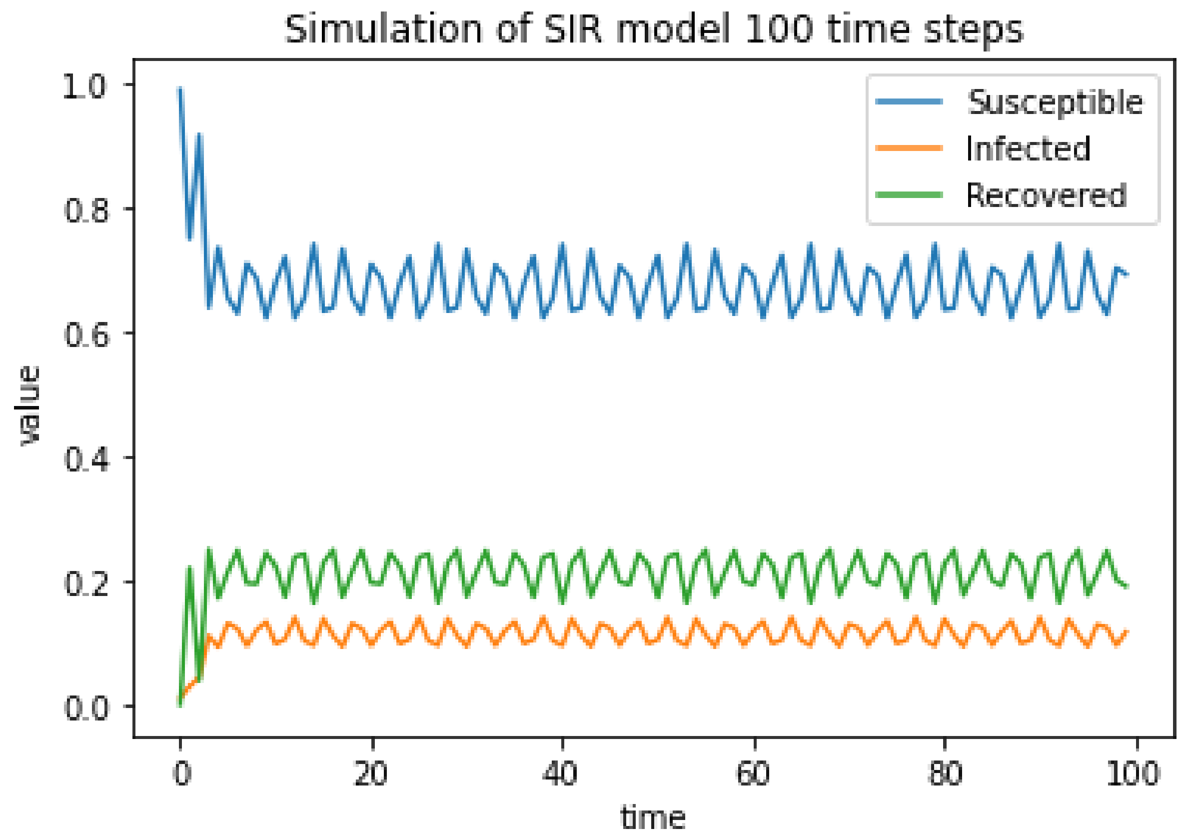
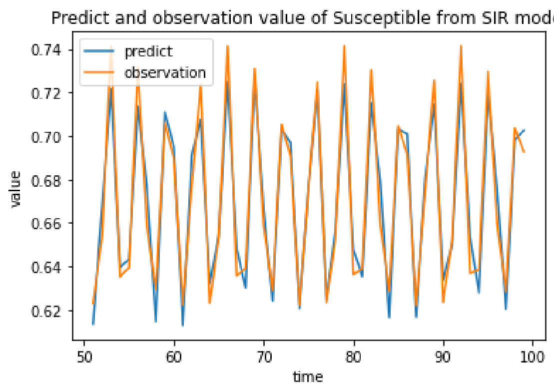
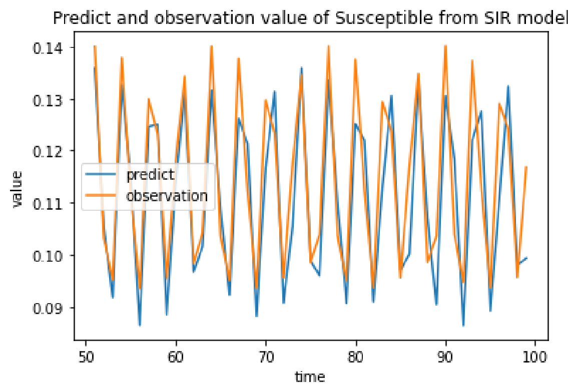


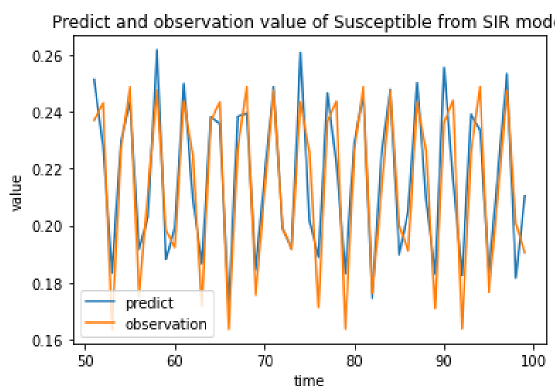
Figure 7: Simulation of SIR model with 100 time steps



(a)

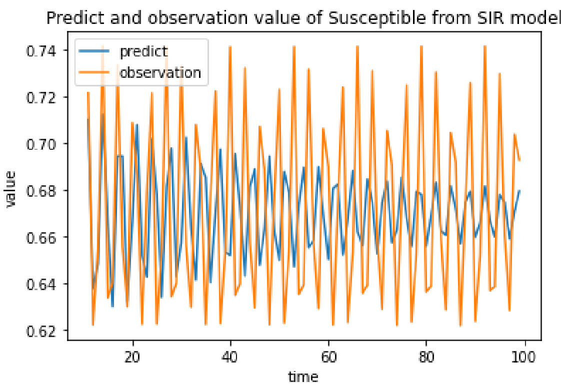


(b)

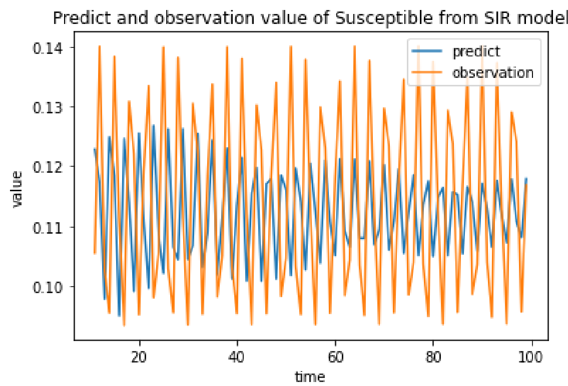


(c)

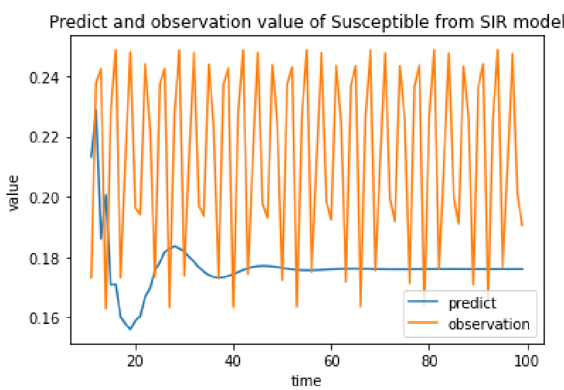
Figure 8: Predict result for (a) Susceptible (b) Infected (c) Recovered in the next 50 time steps



(a)



(b)



(c)

Figure 9: Predict result for (a) Susceptible (b) Infected (c) Recovered in the next 90 time steps

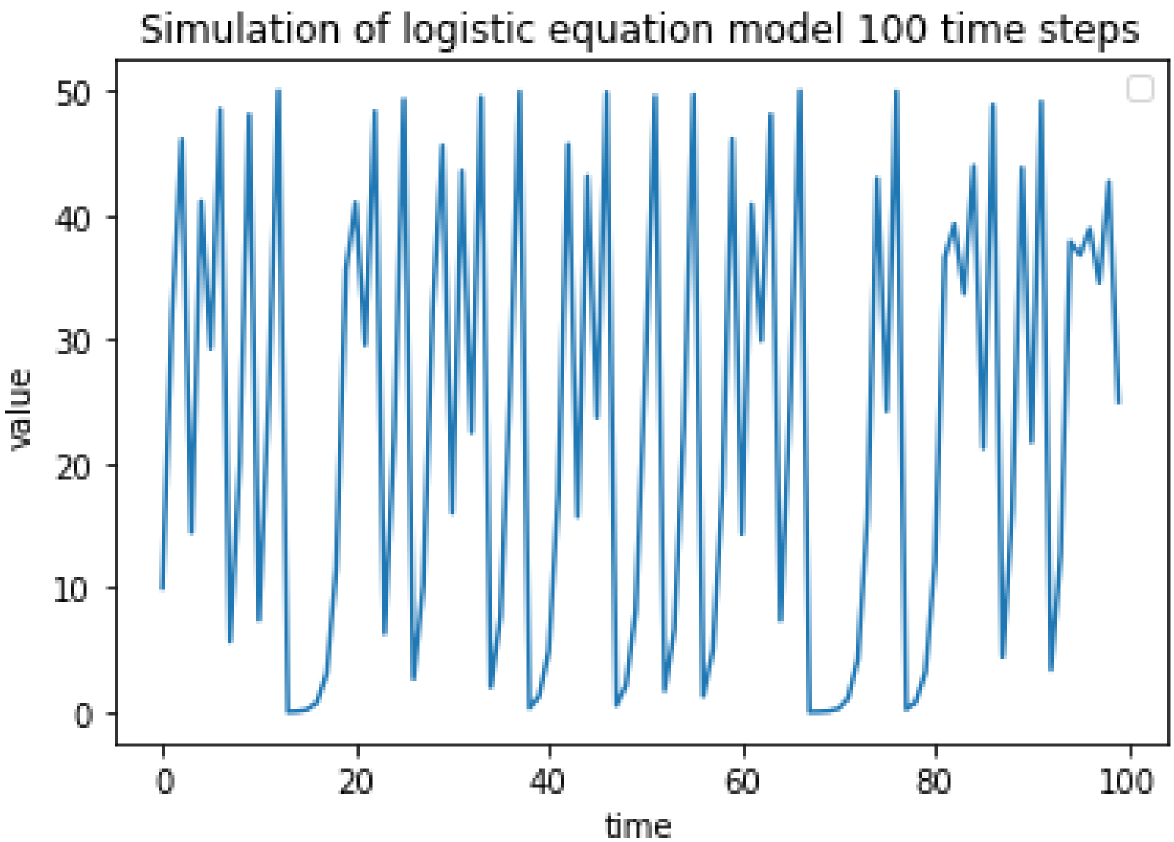


Figure 10: Simulation of Logistic equation with 100 time steps

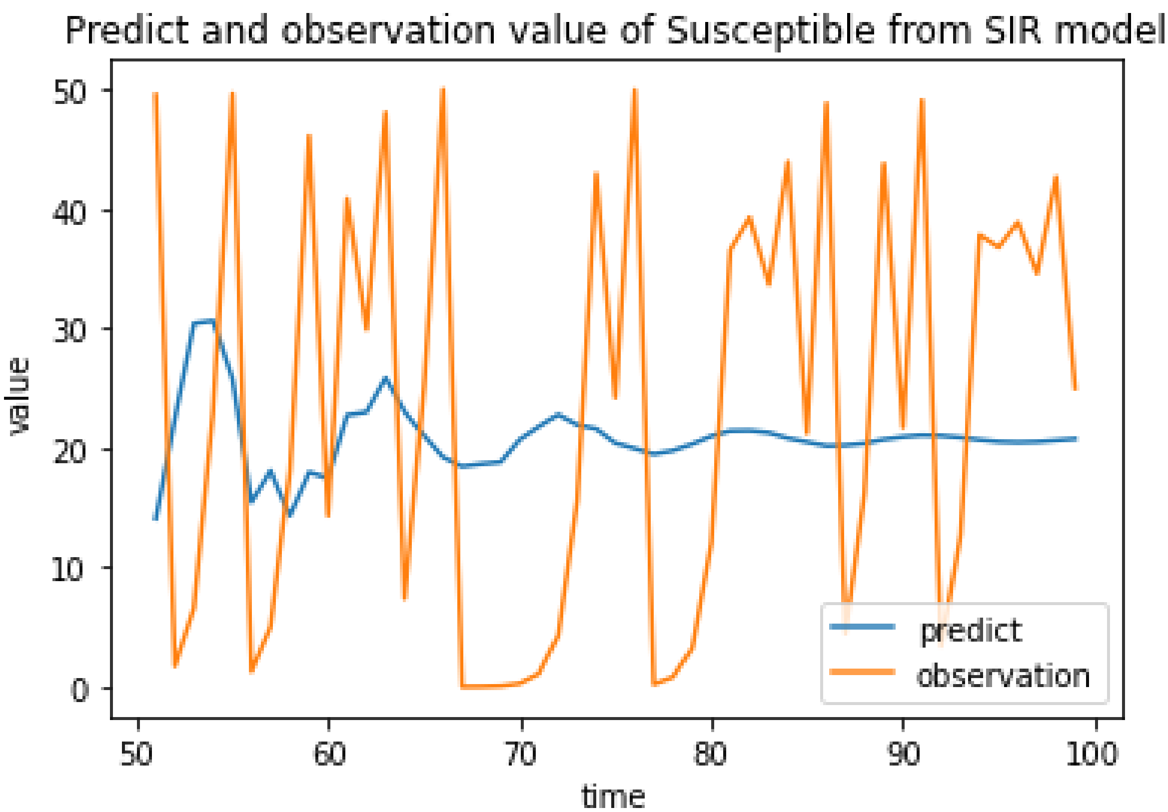


Figure 11: Predict result for logistic equation in the next 50 time steps